challenged by payers and regulatory authorities to develop evidence describing the burden of illness and justifying the payer investment.

HC2
HEALTH CARE EXPENDITURES AND DEPRESSION AMONG ADULTS WITH CANCER

Pan X[1], Sambamoorthi U[2]
[1]West Virginia University, Morgantown, WV, USA, [2]West Virginia University School of Pharmacy, Morgantown, WV, USA

OBJECTIVES: Determine the relationship between health care expenditures and depression in individuals with cancer compared to those with cancer and depression, after controlling for demographic, socio-economic, access to care and other health status variables. **METHODS:** Cross-sectional data on 4766 adults from multiple years (2006, 2007, 2008, and 2009) of the nationally representative household survey, Medical Expenditure Panel Survey (MEPS) were used. Cancer and depression was identified from Medical conditions file. Dependent variables consisted of total, inpatient, outpatient, emergency room, prescription drugs and other expenditures. OLS on logged dollars and generalized linear models with log-link were performed. All analyses accounted for the complex survey design of the MEPS. **RESULTS:** Overall, 14% of individuals with cancer reported having depression. Among individuals with cancer and depression the average health care expenditures were $18,401compared to $12,091 among those without depression. After adjusting for demographic, socio-economic, access to care and other health status variables, those with depression had about 32% greater total expenditures compared to those without depression. Expenditures for every type were higher among individuals with depression compared to those without depression. Individuals with cancer and depression were more significantly more likely to use emergency rooms (AOR = 1.46) and prescription drugs (AOR = 3.56) compared to their counterparts without depression. **CONCLUSIONS:** Among adults with cancer, those with depression had higher health care utilization and expenditures compared to those without depression. Policy efforts to reduce excess health care expenditures associated with depression may include screening for depressive symptoms and preventing major depression, timely depression treatment once depression is detected.

HC3
DISCREPANCIES BETWEEN FDA APPROVAL AND CMS COVERAGE FOR DRUGS AND DEVICES

Chambers JD, Neumann PJ
Tufts Medical Center, Boston, MA, USA

OBJECTIVES: Following FDA approval, medical technology must still gain Centers for Medicare and Medicaid Services (CMS) coverage before Medicare reimbursement. However, the two agencies use approval processes based on different evidentiary standards. We identified the type and nature of discrepancies between FDA approval and CMS national coverage determinations (NCDs) for drugs and devices. **METHODS:** We used the Tufts Medical Center NCD database, which contains detailed information on 165 NCDs since 1999. For each device or Part B drug considered in an NCD (1999-2011) (n=69), we searched the FDA website to identify the approved indication. We classified CMS coverage as: *more restrictive* than FDA approval, i.e., conditions were placed on coverage beyond the FDA-approved label; *equivalent* to FDA approval; or *less restrictive* than FDA approval, i.e., CMS covers off-label indications. Further, we categorized conditions placed on CMS coverage as: "*patient-related*", e.g., restricted to patients with certain comorbidities or characteristics; "*place in therapy*", e.g., tied to use as second-line therapy; or "*technology-related*", e.g., restricted to a particular application of the drug or device. **RESULTS:** CMS has covered FDA-approved drugs or devices taken through the Medicare NCD process in 80% of cases (55/69). For CMS covered drugs and devices (n=55), coverage was more restrictive in 32 cases (58%), equivalent to FDA approval/clearance in 16 (29%), and less restrictive in seven (13%). Most common coverage restrictions were *patient-related* (78%), e.g., laparoscopic gastric banding to treat obesity is covered for patients suffering from an obesity-related comorbidity, and *place in therapy* (38%), e.g., coverage for extracorporeal immunoadsorption is covered for rheumatoid arthritis patients who have failed three disease-modifying antirheumatic drugs (DMARDs). In roughly one third of cases, CMS placed multiple restrictions on coverage. **CONCLUSIONS:** CMS coverage determinations are generally more restrictive than corresponding FDA approval. CMS often restricts coverage to patients with the most severe disease.

HC4
HOSPITALIZATION COSTS AND OUTCOMES AMONG ELDERLY CANCER PATIENTS IN THE UNITED STATES

Khanna R[1], Jariwala K[2], Bentley JP[1], Patel A[3]
[1]University of Mississippi, University, MS, USA, [2]University of Mississippi, Oxford, MS, USA, [3]Medical Marketing Economics, LLC, Oxford, MS, USA

OBJECTIVES: To assess the patient-, hospital-, and discharge-level characteristics, and hospitalization rates among elderly patients with cancer in the United States (US). Hospitalization outcomes (length of stay [LOS], total charges, and mortality) among elderly patients with cancer were also studied. **METHODS:** A cross-sectional descriptive analysis of the 2009 Healthcare Cost and Utilization Project (HCUP) database was conducted. Patients were identified based on diagnosis (any-listed) of cancer using Clinical Classification Software (CCS). A control group of patients without cancer were identified by matching on age and gender (1:2 case-control). Analyses were conducted using PROC SURVEY procedures in SAS v9.2. **RESULTS:** In 2009, a total of 3,325,174 (weighted) hospitalizations occurred among elderly patients with cancer in the US. Elderly cancer patients had higher total hospital charges ($39,406 vs. $37,756), longer LOS (5.7 days vs. 5.4 days), and higher mortality (4.8% vs. 3.6%) as compared to those without cancer. A greater proportion of hospitalizations among cancer patients occurred in teaching hospitals (44.1% vs. 38.9%; p<0.001). In terms of location, a greater proportion of hospitalizations for cancer patients occurred in hospitals located in urban areas in comparison to those without cancer (88.1% vs. 84.7%; p<0.001). Total charges for hospitalizations among elderly patients with prostate (average LOS=4.9 days), lung (average LOS=6.1 days), and breast cancer (average LOS=4.9 days) were roughly $19.2, $16.1, and $16.0 billion, respectively. Mortality rates during hospitalization were the highest for those with pancreatic (10%), liver (9.7%), and lung cancer (8.7%). **CONCLUSIONS:** Elderly patients with cancer had significantly greater hospitalization burden as compared to those without cancer. Hospital mortality rates were the highest for elderly patients with pancreatic, liver, and lung cancer, respectively.

PODIUM SESSION I: HEALTH TECHNOLOGY ASSESSMENT STUDIES

HT1
INTER-COUNTRY VARIABILITY IN COVERAGE DECISIONS FOR ORPHAN DRUGS: CRITERIA DRIVING HTA RECOMMENDATIONS IN SIX COUNTRIES

Nicod E, Kanavos P
London School of Economics and Political Science, London, UK

Inter-country variability in access to orphan drugs across countries has been highlighted in a number of studies. Understanding the reasons driving coverage decisions is a way forward in identifying areas where HTA methods may be improved. **OBJECTIVES:** Objectives are three-fold: a) to establish a methodological framework enabling to systematically compare HTA processes across countries; b) to identify the criteria driving HTA recommendations for a sample of orphan drugs, and; c) to understand the reasons for diverging recommendations and propose ways to minimize these. **METHODS:** All common orphan drug-indication pairs appraised in six countries (England, Scotland, France, Sweden, Canada and Australia) between 2001 and 2012 were selected. Agreement levels in HTA outcomes between countries were measured using Cohen's kappa scores. Thematic analysis, by creating an NVivo-9 coding manual, was conducted to systematically compare each compound. Reasons for diverging HTA outcomes were differentiated based on whether they are a consequence of country-specific considerations or of the HTA process, and ranked by frequency of occurrences. **RESULTS:** Fourteen orphan drug-indication pairs were retrieved. Agreement in HTA outcomes was poor (k = [-0.5; 0.3]). Eight drug-indication pairs appraised by at least four HTA bodies were analyzed, five of which received diverging outcomes. Preliminary results suggest that in four of five cases, reasons for diverging recommendations were a consequence of the HTA process. Examples of non-homogeneous assessments include: lack of appropriate primary endpoint, lack of long-term data, evidence not reflecting clinical practice, orphan status or unmet clinical need. **CONCLUSIONS:** Preliminary results identify the criteria driving the assessments and reasons why they result in diverging HTA outcomes, enabling a better understanding of these processes by elucidating the expectations and value judgments from HTA bodies, particularly on the orphan status, and identifying areas where more consensus on what constitutes appropriate HTA methodologies is needed. Final results will quantify these criteria in a systematic manner.

HT2
AGENCY AGREEMENT IN HEALTH TECHNOLOGY ASSESSMENT REIMBURSEMENT DECISIONS

Jaksa A, Demb J, Ho YS, Daniel K
Context Matters, Inc., New York, NY, USA

OBJECTIVES: HTA agencies often review the same drugs for the same/similar indications. How often do agencies agree on their reimbursement decisions? Previous research has compared reimbursement recommendations (for the same drugs) for a limited number of agencies, but studies have rarely focused on more than 2 agencies. We collect and analyze a large number of health technology assessments from several countries to explore how often the agencies agree on their reimbursement decisions. **METHODS:** The data covered five agencies that make reimbursement decisions: NICE, SMC, PBAC, HAS and CADTH's Common Drug Review. Our analysis only included decisions for drugs that were reviewed by at least two agencies. If a drug was reviewed multiple times by an agency for the same indication (i.e. resubmissions or updates) we used the most recent review for the analysis. A total of 78 drugs were reviewed by at least 2 agencies, producing a total of 195 reviews. **RESULTS:** There was generally a high level of agreement between all pairs of agencies, ranging from 56% (PBAC; CADTH) to 91% (NICE; HAS). It is important to note that within the sample of drugs reviewed, all agencies issued positive recommendations at very high rates – all but CADTH issued positive recommendations for more than 80% of the drugs reviewed. This fact alone would produce high levels of agreement, even if agencies' recommendations were statistically independent. Actual agreement rates observed were close to those implied by independence. **CONCLUSIONS:** Agencies agree on their reimbursement decisions quite often, but at rates close to those implied by their high overall positive recommendation rates alone. Future research will focus on identifying the determinants of agencies' high rates of agreement.

HT3
ISSUES IN THE SELECTION OF COMPARATORS FOR REGULATORY AND HTA SUBMISSIONS

DeLozier AM[1], Murray JF[1], Klopchin MN[1], Johnston JA[1], Grainger D[2]
[1]Eli Lilly and Company, Indianapolis, IN, USA, [2]Eli Lilly and Company USA /Australia, West Ryde, Australia

OBJECTIVES: In developing new medicines, pharmaceutical companies face the complex and challenging task of choosing comparators, in the absence of clear and consistent guidance, in order to produce evidence that best meets diverse stakeholder needs. This lack of clear and consistent guidance has led to several high-profile cases in which HTA bodies rejected products receiving regulatory approval due to a lack of evidence relative to desired comparators. Our objective was to identify key issues in comparator selection and present case studies highlighting these issues. **METHODS:** We conducted a comprehensive search of the published and grey literature addressing comparator selection, including formal guidelines from major regulatory, HTA, and professional organizations, as well as case studies and published opinions. Three authors independently and qualitatively assessed the identified literature, and the final list of issues was agreed upon through team discussion and consensus. **RESULTS:** Key challenges in the selection of appropriate comparators include: 1) lack of a standard definition for "standard of care," 2) concerns regarding off-label usage; 3) changes in requirements due to market dynamics; and 4) RCT design issues. We explore several of these challenges further with case studies. **CONCLUSIONS:** Due to challenges inherent in meeting diverse stakeholder needs, the selection of comparators is a complex and multi-faceted problem. The challenges we identified highlight the need for more uniformity and clarity among existing standards. In the context of ongoing external standard setting and harmonization efforts, organizations seeking to provide guidance around comparator selection should consider the issues described herein and work towards the common goal of a more transparent, harmonized process that optimizes the production of evidence that best meets the needs of all stakeholders.

HT4

HEALTH TECHNOLOGY ASSESSMENT LANDSCAPE OF BIOSIMILARS – APPROACHES TO EVALUATIONS AND RESULTS

Andreykiv M[1], Dahal D[2]
[1]*Quintiles, Hoofddorp, The Netherlands,* [2]*Quintiles, Hawthorne, NY, USA*

OBJECTIVES: Gain insight into how HTA agencies evaluate biosimilar products and understand implications for future biosimilars. **METHODS:** Step I: Manually search 60 health care agencies' websites for reports evaluating biosimilars. Step II: Categorize reports by HTA type, scope and outcome. Step III: Using a standardized set of categorical criteria, investigate HTA agencies' approaches to biosimilar treatment recommendations, across multiple therapeutic areas. **RESULTS:** We identified a total of 47 HTA reports evaluating biosimilars, of which 38 are Single Technology Assessments (STA) and 9 are class reviews or clinical guidelines. Except for one STA by AMWSG, all recommendations were positive and no agency recognized substantial differences in clinical efficacy or safety between original and biosimilar products. A major factor influencing the negative AMWSG recommendation was limitations in the economic model provided by the sponsor company. After reviewing HTA reports on biosimilars, it is apparent that different countries apply different strategies to evaluating biosimilars – In France (13 STAs), Scotland (7 STAs) and Sweden (6 STAs), each biosimilar was reviewed individually; in Netherlands, CVZ published a "Preference policy for biologics" classifying biological products as "similar" and therefore therapeutically interchangeable. KCE (Belgian HTA body) chose this same approach and is currently studying the advantages and barriers of using biosimilars in Belgium. **CONCLUSIONS:** HTA agencies that have reviewed each biosimilar product individually, assessing all the available data through the full HTA process, have acknowledged the comparable safety and efficacy of biosimilar products with the original molecule. However, some individual restrictions were made in terms of prescribing patterns or formulation type limitations. Although the standard HTA path for biosimilars is not fully defined, we can conclude from this evaluation that HTA agencies generally treat biosimilars as equivalent in safety and efficacy to originator products.

PODIUM SESSION I: RESEARCH ON METHODS – PATIENT-REPORTED OUTCOMES STUDIES

PO1

APPLICATION OF ITEM RESPONSE THEORY IN VALIDATING THE MORISKY MEDICATION ADHERENCE SCALE IN PATIENTS WITH HYPERTENSION

Gupta S[1], Goren A[2]
[1]*Kantar Health, Princeton, NJ, USA,* [2]*Kantar Health, New York, NY, USA*

OBJECTIVES: The Morisky Medication Adherence Scale (MMAS) has been validated in patients with hypertension using classic test theory analyses. Item Response Theory (IRT) methods are an increasingly popular, sophisticated means of validation in health outcomes research. The current study examined psychometric properties of the eight-item MMAS (MMAS-8) among hypertensive patients, using IRT. **METHODS:** Data were used from the U.S. 2012 National Health and Wellness Survey (NHWS; N=71,141), a web-based survey of patients' demographics, health care attitudes and health outcomes. Respondents reporting physician diagnosis of hypertension and currently taking a prescription medication for hypertension were included. Reliability and validity of the MMAS-8 were examined using both classic test theory (Cronbach's alpha) and IRT (two-parameter graded response model) analyses. **RESULTS:** Among 16,680 patients with hypertension, 54.21% were male, mean age was 60.88 (SD = 12.74) and years since diagnosis were 13.40 (SD=10.82). Non-adherence behavior frequencies varied across items; "do you have difficulty remembering to take all your medicine?" (28.92% "yes"), "do you sometimes forget to take your medicine?" (21.97%), "do you feel hassled about sticking to your treatment plan?" (16.21%) and "cut back or stopped taking your medicine because you felt worse when you took it?" (4.76%). The MMAS-8 was adequately reliable (Cronbach's a=0.71); however, improvement was seen with removal of the "did you take all your medicine yesterday?" item (a=0.73). The eight items varied in their discrimination (range: 0.53–3.09), and item location parameters reflected a large, high range of non-adherence (0.68–4.79), indicating that the item set discriminates best at higher levels of non-adherence. **CONCLUSIONS:** The MMAS-8 has adequate psychometric properties for evaluating non-adherence in patients with hypertension, with room for improvement by eliminating one item and adding items that discriminate at lower levels of non-adherence. IRT is a helpful means for evaluating the reliability and validity of the MMAS-8.

PO2

PSYCHOMETRIC PERFORMANCE OF THE NEI VFQ-25: RASCH ANALYSIS OF THE NEI VFQ-25 AS A MEASURE OF PATIENT-REPORTED VISUAL FUNCTION ACROSS FOUR RETINAL DISEASES

Petrillo J[1], Ferreira A[2], Cano S[3], Bressler NM[4], Lamoureux E[5]
[1]*Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA,* [2]*Novartis Pharma AG, Basel, Switzerland,* [3]*ScaleReport, Stotfold, UK,* [4]*John Hopkins University School of Medicine, Baltimore, MD, USA,* [5]*University of Melbourne, Melbourne, Australia*

OBJECTIVES: The 25-item National Eye Institute Visual Function Questionnaire (NEI VFQ-25) is widely used to assess patient-reported visual functioning. Previously, the reliability and validity of the NEI VFQ-25 has been assessed in patients with neovascular age-related macular degeneration (AMD). Its measurement performance in other ocular indications, and in large pooled datasets, remains unclear. This study aimed to evaluate the psychometric properties of the NEI VFQ-25 in a large sample of patients across multiple retinal diseases. **METHODS:** Dataset included pooled baseline NEI VFQ-25 data from six clinical trials in diabetic macular edema (DME), macular edema from branch and central retinal vein occlusion (RVO), neovascular AMD, and myopic choroidal neovascularisation. Rasch analysis was conducted by assessing item fit validity, threshold targeting, item dependency, reliability and stability. **RESULTS:** Measurement performance was evaluated for 2487 person measurements (mean age: 64±9 (SD) years; gender: 53% male). Mixed psychometric properties of the NEI VFQ-25 were identified. Key strengths identified included a high Person Separation Index of 0.93 (suggesting good reliability), predominantly low residual correlations (suggesting minimal local dependency between items), and no statistically significant Differential Item Functioning (suggesting stability of scoring function across country, study and gender). Some potential psychometric limitations of the NEI VFQ-25 included disordered thresholds for 15 of the 25 items, poor fit according to the Rasch model for several items, and mismatched distributions of person and item threshold locations. These limitations suggest sub-optimal item targeting, potential problems with conceptual clarity, lack of discrimination across some response options and a consistent ceiling effect across all indications. **CONCLUSIONS:** Rasch analysis of NEI VFQ-25 identified many positive characteristics of this widely used instrument. Several suboptimal characteristics, however, were also identified, suggesting that changes to the NEI VFQ-25 could be considered to improve its psychometric performance in clinical trials of retinal diseases.

PO3

THE PAIN ASSESSMENT FOR LOWER BACK IMPACTS (PAL-I): QUALITATIVE DEVELOPMENT AND COGNITIVE EVALUATION OF A NEW PATIENT REPORTED OUTCOME MEASURE FOR THE ASSESSMENT OF IMPACTS OF LOW BACK PAIN

Ramasamy A[1], Blum SI[1], McCarrier KP[2], Quintanar-Solares M[2], Bushnell DM[2], Liedgens H[3], Martin ML[2], Argoff CE[4], Patrick DL[5], Wallace MS[6], Freynhagen R[7]
[1]*Forest Research Institute, Jersey City, NJ, USA,* [2]*Health Research Associates, Inc., Seattle, WA, USA,* [3]*Grünenthal GmbH, Aachen, Germany,* [4]*Albany Medical College, Albany, NY, USA,* [5]*University of Washington, Seattle, WA, USA,* [6]*University of California, San Diego, La Jolla, CA, USA,* [7]*Benedictus Krankenhaus Tutzing, Tutzing, Germany*

OBJECTIVES: To develop a patient-reported outcome (PRO) measure to assess the key functional and lifestyle impacts of chronic low back pain (cLBP) through qualitative concept elicitation (CE) and cognitive interviews. **METHODS:** Adult patients (18-80 years) with clinical diagnoses of cLBP of non-malignant origin experiencing moderate to severe levels of pain intensity were recruited in the U.S., UK, and Germany. Trained interviewers conducted CE and cognitive interview sessions with participating patients using semi-structured interview guides. The CE interviews elicited spontaneous reports of symptoms and associated impacts, followed by further probed exploration to confirm concepts. Cognitive interviews evaluated the degree of patient comprehension of the items in the draft PRO measure. All interview sessions were audio recorded and transcribed. The CE interviews were coded for qualitative content analysis using Atlas.ti, and cognitive interview transcripts were summarized in cognitive report tables. **RESULTS:** Forty-three CE interviews were conducted (mean age: 48.6±13.0, 53.5% female, 74.4% -White/Caucasian). Mean pain NRS score was 6.7±1.3. A total of 2,220 impact expressions were derived from the transcripts, representing 47 different impact concepts. Data from CE interviews was considered alongside existing measures, published literature and expert opinion to develop a 15-item draft instrument. Thirty additional patients participated in four waves of cognitive interviews, during which two items were removed and others were substantially modified to create the Pain Assessment for Low Back Impacts (PAL-I). **CONCLUSIONS:** The PAL-I is a 13-item PRO measure for assessing lifestyle and functional impacts of cLBP that has been developed through direct qualitative patient involvement in accordance with the FDA's PRO Guidance and scientific best practices. Cognitive interviews have provided evidence from patients that the measure is comprehensive, relevant to their cLBP experience, comprehendible, and easy to complete. The measure will be further tested in additional qualitative and quantitative studies to evaluate its measurement properties.